SCENE CHANGE IDENTIFICATION DURING ENCODING OF COMPRESSED VIDEO

Inventors: Vikrant Kasarabada, Vaidyanath Mani, and Haitao Guo

Field of Invention

[0001] The present invention relates generally to scene change detection and bit rate control in video encoding, and more particularly, to scene change identification and bit rate control during the encoding of compressed video images.

Background of the Invention

[0002] Scene change detection in digitized video is the process of identifying changes in the video data that are likely to represent semantically meaningful changes in the video content, typically associated with a change in the shot or scene. Scene change detection is useful particularly for video indexing and archiving, and allows for user to "browse" video. Conventional scene change detection is typically applied during the playback of video. Numerous algorithms have been created to identify scene changes, relying variously on histogram analysis of color, luminance, and other image characteristics, and in some cases on analysis of motion vectors. These approaches have been likewise extended to identify scene changes in compressed video content, such as MPEG-2 video. However, these approaches require the computationally expensive process of first decoding the compressed video prior to do the scene change analysis, in order to obtain the necessary statistics and data, such as motion vectors, color data, and the like.

[0003] Some researchers have explored using motion vector information created during the MPEG encoding process, relying on the amount and magnitude of motion vectors, as well as mean square prediction error, to identify scene changes relative to a last reference frame. This approach however may falsely identify scene changes where there has been gradually changing image content in an otherwise continuous scene. In addition, these methods do not advantageously adjust the bit rate used for image encoding, but instead attempt to improve image quality to forcing a change in the frame type.

[0004] Accordingly its desirable to provide an improved scene identification process during the encoding of compress video. It is further desirable to use the scene change information to modify the bit rate for improved image quality.

Summary of the Invention

[0005] The present invention includes a scene change identification process that operates during the encoding of compressed video, and that provides for flexible control of the bit rates used during the encoding process. Generally, an uncompressed image is received in an encoder or encoding process. The uncompressed image comprises a macroblocks, and will have a frame type, such as an I frame, a P frame, or a B frame, as is known in the art. Each macroblock will also have a type, corresponding to one of these types, resulting from a motion compensation process. From the distribution of the types of the macroblocks, a determination is made as to whether the image corresponds to a scene change. In one embodiment, the determination of whether the image corresponds to a scene change is made based on various percentages of predicted motion macroblocks relative to a total number of macroblocks. The percentage of motion blocks is compared with a threshold to determine whether the frame represents a scene change. Based on this determination and the type of the frame, a scene change

is determined. Preferably, when a scene change is detected, the encoder allocates a greater number of bits to the frame, without changing the frame type. The frame is then encoded based on the rate control parameter. The encoder can also generate a side information file that contains the scene change information to create a scene table of contents.

Brief Description of the Drawings

[0006] FIG. 1 is a block diagram of a video encoder according to one embodiment of the invention.

Detailed Description

[0007] Referring now to FIG. 1, there is shown a block diagram of an encoder 101 according to one embodiment of the present invention. The encoder 101 generally operates according to principles of an MPEG-1 or MPEG-2 video encoder as described in ISO/IEC 11172 and ISO/IEC 13818 or similar hybrid DPCM/DCT encoder, with the additional features and structures as further described below. The encoder 101 is used to encode a video sequence that comprises a plurality of uncompressed images 10. These images are input into the encoder 101 along with side information that describes the bit-rate and the group of pictures pattern to be used in the encoding. The encoder then determines the position of an image in the group of pictures. The image 10 data comprises a plurality of macroblocks, each macroblock having either 4:x:x sample format of luminance and chrominance data, depending on the implementation.

[0008] The type of image (equivalently "frame" or "picture") will be one of an intra-pictures (I), a forward predicted picture (P), or a bi-directional predicted (B) picture. Intra-pictures (I-pictures) are coded without reference to other pictures contained in the video sequence. Inter-frame predicted pictures (P-pictures) are coded with reference to the nearest previously coded I-picture or P-picture, usually

incorporating motion compensation to increase coding efficiency. Bi-directionally predicted (B-pictures) use both past and future frames as references. To achieve high compression, motion compensation can be employed based on the nearest past and future P-pictures or I-pictures. B-pictures themselves are never used as references.

[0009] The encoder 101 uses different encoding processes depending on the type of the image. P and B type frames are encoded as follows. The image macroblocks are input into both subtractor 40 and motion estimator 20. The motion estimator 20 compares each of these current image's macroblocks with macroblocks in a previously stored reference picture or pictures. For each current macroblock, the estimator 40 finds a macroblock in a reference picture that most closely matches the current macroblock. The motion estimator 40 then calculates a motion vector that represents the horizontal and vertical displacement from the current macroblock being encoded to the matching macroblock in the reference picture. When completed for all of the macroblocks in the current image, the result is a set of motion vectors corresponding to the macroblocks of the current image. Each macroblock will be also coded as either a forward predicted (P), backward predicted (B), intra (I), or skip (S) block, depending on the motion vector information for the block, if any, and the frame type. For a B frame, motion predicted blocks will be either P or B blocks, depending on the reference picture used for the particular macroblock; otherwise, the blocks will be I or S blocks. For a P frame, the blocks will be either P, S, or I blocks. These motion vectors data and block information are passed to the motion compensation stage 25, which applies them to the reference image(s) to create a motion compensated image 30.

[0010] The motion compensated image 30 is then subtracted from the original image 10 by subtractor 40 to produce a set of error prediction or residual signals for each macroblock (in practice this step takes place on macroblock by macroblock basis directly following motion estimation). This error prediction signal represents the difference between the predicted image and the original image 10 being encoded. In the case that

the original image 10 is a B- or P-picture, the motion compensated image 30 is an estimation of the original image 10. In the case that the original image 10 is an I-picture, then the motion compensated image 30 will have all pixel values being equal to zero, and the quantized DCT coefficients represent transformed pixel values rather than residual values as was the case for P and B pictures. For I frames, the quantized coefficients are used to reconstruct the I frame as a reference image via inverse quantizer 100 and inverse DCT 110. The reconstructed image is store in frame buffer 115.

[0011] In accordance with the present invention, the macroblock type data from the motion estimator 20 is input into scene change detector 50. The scene change detector 50 determines whether there has been a scene change at the current image based on the distribution of the macroblock types, such as the I, P, B, and S blocks, and the frame type of the current frame. In one embodiment, the scene change detector 50 identifies a scene change by comparing the percentages of P, B, or I blocks in the image to one or more thresholds. Preferably, the particular threshold comparison depends on the picture type of the current frame, since the distribution of block types in a scene change tends to differ depending on the frame type. If the percentage of P, B or I blocks exceeds the threshold amount(s), then the image is deemed to be a scene change. (Those of skill in the art appreciate that the comparisons can be equivalently implemented so that a scene change is determined if the percentages do not exceed the thresholds).

[0012] Having determined whether the current image is a scene change, the scene change detector 50 provides a scene change flag to the quantizer 70 to indicate a scene change. The quantizer 70 responds to the scene change flag by changing the increasing the number of bits used to quantize the image, thereby increasing the quality of the encoded image. This may be done either directly, by changing the quantization step size, or indirectly by changing adjusting the rate controller internal to the quantizer 70. The details of the quantization control are further described below.

[0013] In one embodiment, the scene change detector 50 operates as follows. As noted above, there are three types of frames or pictures that will be received by the encoder: I, P, and B frames. If the current frame is an I frame, and a scene change (as would be perceived by a viewer) has in fact occurred here, there is no need for the scene change detector 50 to instruct the quantizer 70 to allocate additional bits during quantization. This is because the I frame will already be entirely intra-coded, and will provide sufficient image quality. In one embodiment in which the scene change detector 50 builds a scene table of contents 130, the scene change detector 50 generates index information that identifies the current image as a scene change, for example by outputting the current group of pictures (GOP) index, and the frame number of the current frame within the current GOP. The scene TOC 103 can be transmitted to, subsequently decoded by a complementary decoder to provide the viewer with access to a scene change table of contents or the like.

[0014] The second case is where the current frame is a B frames, which will have two reference frames, one forward and one behind. If a scene change occurs at a B frame, then the motion vectors for that frame should mostly point in one direction, either forward, or backward, and hence the distribution of P and B blocks will indicate a scene change.

[0019] The scene change detector 50 uses a threshold-based test to detect this situation, and to ignore the false positives (that is, instances that are not actual scene changes). In one embodiment, the tests and thresholds are determined as follows. A total number macroblocks (M) is set. M can be a fixed, based on the total number of macroblocks given the frame size. Alternatively, M can be set to the total number of motion predicted blocks (including forward and backward predicted); this is the same as the total number of macroblocks in the frame minus S and I blocks. The scene change detector 50 then calculates various ratios:

[0020] Percent Forward Predicted (PF) = number of forward predicted blocks/M

[0021] Percent Backward Predicted (PB)= number of backward predicted blocks/M

[0022] In one embodiment, the scene change threshold for these percentages is set at 70%. Accordingly, if either PF or PB is greater than .70, then a scene change is declared. Where I and S blocks are not included in M, the scene change threshold is correspondingly decreased, with the appropriate percentages based on sample video data. Also, an optional threshold test is the percentage of I blocks relative to M. This test operates on the converse principle to the above tests, and if the percentage is below a selected threshold, then a scene change is declared. The selection of which type of test to use is a matter of implementation preference, as both types of tests yield essentially the same result.

[0023] The third case is where the current frame is a P frame. As described above, a P frame is encoded based on a prior I frame. The blocks in the P frame then will be distributed amount I, S, and P blocks. If a scene change has not occurred at the current P frame, then there should be a high proportion of P blocks, since the will be well predicted by the prior I frame. On the other hand, a scene change at the P frame would correspond to a significant amount of new material in the frame, as compared to a prior reference frame. This means that for such a scene change, the motion estimator 20 should have unable to find a significant number of matches for the P frame from the blocks of prior reference frame. As a result, in a P frame, a scene change correlates to a relatively high number of I blocks in the frame (since relatively few blocks were motion predicted).

[0024] More specifically, the scene change detector 50 can implement the foregoing logic in a variety of ways. In one embodiment, the scene change detector 50 determines the percentage of I blocks (PI) relative to M. If PI is greater than 65%, then a scene change is declared. Alternatively, the scene change detector 50 can compute the ratio of I blocks to P blocks, and declare a scene change when this ratio is greater than

about 2 (that is, more than about two thirds of the frame are I blocks. Those of skill in the art will appreciate the variety of ways in which statistics of the distribution of blocks types can be computed and selected in view of the teachings of the present invention.

[0025] Regardless of how the scene change detector 50 identifies a scene change in either a P or B frame, the scene change detector 50 provides a scene change flag to the quantizer 70. This is done without changing the type of the current image to an I frame (from either a P frame or a B frame) as is done in conventional approaches to which use scene change information to improve encoding. Instead, the present invention maintains the frame type of the current image, and instead increases the number of bits used to encode the image, thereby directly improving the image quality.

[0026] Optionally, the scene change detector 50 further generates index information that identifies the current P or B frame as a scene change, by outputting the current group of pictures (GOP) index, and the frame number of the current frame within the current GOP, and updating the scene TOC 130.

[0027] The residual signals from subtractor 40 are transformed from the spatial domain by a two dimensional DCT 60, to produce a set of DCT coefficients for each residual macroblock. The DCT coefficients of the residual macroblocks are quantized by quantizer 70 to reduce the number of bits needed to represent each coefficient. The quantizer 70 uses a bit rate control algorithm that adjusts the bit rate based on the image type and the scene change flag received from the scene change detector 50. For P type pictures, the quantized coefficients are passed back through an inverse quantizer 100 and inverse DCT 110 to creates a reconstructed image which is stored in frame buffer 115, and made available as input into the motion estimator 20 and motion compensation 30 as needed. The reconstructed image is equal to the image, which will be found at the receiver side, if no transmission errors have occurred.

[0028] As shown by the control signal path from the buffer 90 to the quantizer 70, the type of the picture and the amount of data in the buffer 90 control the

quantization. In a conventional encoder, if the amount of data in the buffer increases (e.g., when the buffer is holding an image that include significant amounts of intra encoded data), the bit rate control algorithm would increase the quantization step size, resulting in a coarser quantization, which would to decrease the image quality of the current image, even if the current image corresponds to a scene change. The encoder 101 of the present invention however does not suffer from this defect, since the scene change detector 50 can influence the degree of quantization based on the presence or absence of a scene change.

[0029] The quantizer 70 operates as follows, in one embodiment. The quantizer 70 maintains two bit rate variables: T is a variable that tracks the total number of bits allocated for current frame in the GOP. R is a variable that tracks the number of bits remaining after encoding each frame in the GOP. A rate control algorithm 75 in the quantizer 70 uses these variables allocate bits to each frame to be encoded. The rate controller 75 will allow the number of bits allocated to a frame to vary by about +/- 2-3% per frame. The ability of the rate control algorithm 75 to stick to the target is a measure of the efficiency of the rate control and encoder.

[0030] The quantizer 70 receives a scene change flag from the scene change detector 50 with each frame; the flag will indicate whether or not the current frame corresponds to a scene change. If there is a scene change, and the frame type is a P or B frame, the rate control algorithm 75 increases either T or R, or both, by an encoding factor X. This gives the quantizer 70 the illusion, so to speak, that there are actually more bits available for quantizing the image than in fact are allocated.

[0031] The value of the encoding factor X ultimately determines the quantization rate (or step size) used by the quantizer 70, and thus the quality of the current image. However, X cannot be arbitrarily set to any amount. If X is too large, then T and R are increased too much, and the result is that the number of bits used for the frame will exceed the target allocation. If X is too small, then it will not help to increase quality of

the image in any significant way, because the quantization step size will not have been sufficiently decreased. In a preferred embodiment, the bit factor X is variable and is proportional to the current value of T, and both T and R are increased by X. In one embodiment, X = 0.125T; thus:

[0032] T=T+0.125T; and

[0033] R=R+0.125T.

[0034] This allows X to be easily encoded as a power of 2, and thus readily bit shifted to increase or decrease the level of quantization.

[0035] More specifically, increasing the value of R by the encoding factor X results in an increase the target bit rate allocation Tp or Tb depending on whether the current frame is a P or B frame. For a P frame , T = Tp and for a B frame T = Tb. Tp and Tb are generally calculated as follows:

$$T_p = \max\left\{\frac{R}{\left(N_p + \frac{N_b K_p X_b}{K_b X_p}\right)}, \frac{bit\_rate}{8 \times picture\_rate}\right\}$$

$$T_b = \max\left\{\frac{R}{\left(N_b + \frac{N_p K_b X_p}{K_p X_b}\right)}, \frac{bit\_rate}{8 \times picture\_rate}\right\}$$

[0036] Where Kp and Kb are "universal" constants dependent on the quantization matrices, with typical values of Kp = 1.0 and Kb = 1.4; N is the number of pictures in the current GOP; Np and Nb are the number of P-pictures and B-pictures remaining in the current GOP in the encoding order; and bit_rate and picture_rate are previously determined factors.

[0037] From the revised target allocations Tb or Tp, a fullness measure dj^p or dj^b is computed, where:

$$d_j^p = d_0^p + B_{j-1} - \left(\frac{T_p \times (j-1)}{MB\_cnt}\right)$$

$$d_j^b = d_0^b + B_{j-1} - \left(\frac{T_b \times (j-1)}{MB\_cnt}\right)$$

[0038] depending on the picture type, where d_0^i, d_0^p, d_0^p are initial fullnesses of virtual buffers, one for each picture type. Bj is the number of bits generated by encoding all macroblocks in the picture up to and including a current macroblock j; MB_cnt is the number of macroblocks in the picture.

[0039] From the fullness measures, the quantizer 70 then computes the quantization parameter Qj for macroblock j as follows:

$$Q_j = \left(\frac{d_j \times 31}{r}\right)$$

[0040] where the "reaction parameter" r is given by

$$r = 2 \times \frac{bit\_rate}{picture\_rate}$$

[0041] and dj is the fullness of the appropriate virtual buffer. Finally, the quantization parameter Qj is used to quantized the DCT coefficients for the current macroblock. Further details of the quantization process may be found in the documentation for Test Model 5 for the MPEG Software Simulation Group, MPEG Doc. No. MPEG 93/457 (April, 1993).

[0042] In summary then, increasing T and R results in an increase in the number of bits allocated to the current type of frame, which in turn decreases the fullness measure for the current frame type as well. This decrease in the fullness measure in turn reduces Q. In an alternative embodiment, T alone is increased by X, which will

also have a similar net effect on reducing Q. Similarly, R alone can be increased. In yet another embodiment, R and T can be increased by differing amounts. In a further embodiment, Q can be reduced directly based on the above equations.

[0043] After the quantization of the DCT coefficients is performed, then the rate control algorithm subtracts X from T and from R, essentially restoring them to their prior values. R is then updated to reflect the actual number of bits used to encode the current frame. The temporary values T and R do not effect the operation of the inverse quantizer 100.

[0044] Following the quantization, the quantized DCT coefficients are entropy encoded in entropy encoder 80 which further reduces the average number of bits per coefficient. The resulting encoded coefficients are combined with the motion vector data and other side information (including an indication of I, P or B picture) and sent to an image buffer 90, which buffers the output bitstream to produce encoded image 120.

[0045] Once the video sequence of uncompressed images 10 is encoded, the overall output will be an encoded, compressed video file, which can also contain the scene TOC 130. The encoder 101 can augment the scene TOC 130 by including a representative image for each of identified scene changes; the representative image is preferably a subsampled version of the identified frame, and can likewise be encoded prior to transmission or storage. Alternatively, the representative frames can be generated on a compliant decoder by subsampling the identified images in the scene TOC 130. These representative images can then be displayed to a viewer, to allow the viewer to jump to the beginning of each scene as desired.